

8th Annual  
Earth System Grid Federation



# THE STATE OF THE EARTH SYSTEM GRID FEDERATION



**LUCA CINQUINI**

NASA JET PROPULSION LABORATORY AND CALIFORNIA INSTITUTE OF TECHNOLOGY

JPL UNLIMITED RELEASE SYSTEM CLEARANCE NUMBER: #

© 2018 CALIFORNIA INSTITUTE OF TECHNOLOGY. GOVERNMENT SPONSORSHIP ACKNOWLEDGED



# Introduction

---

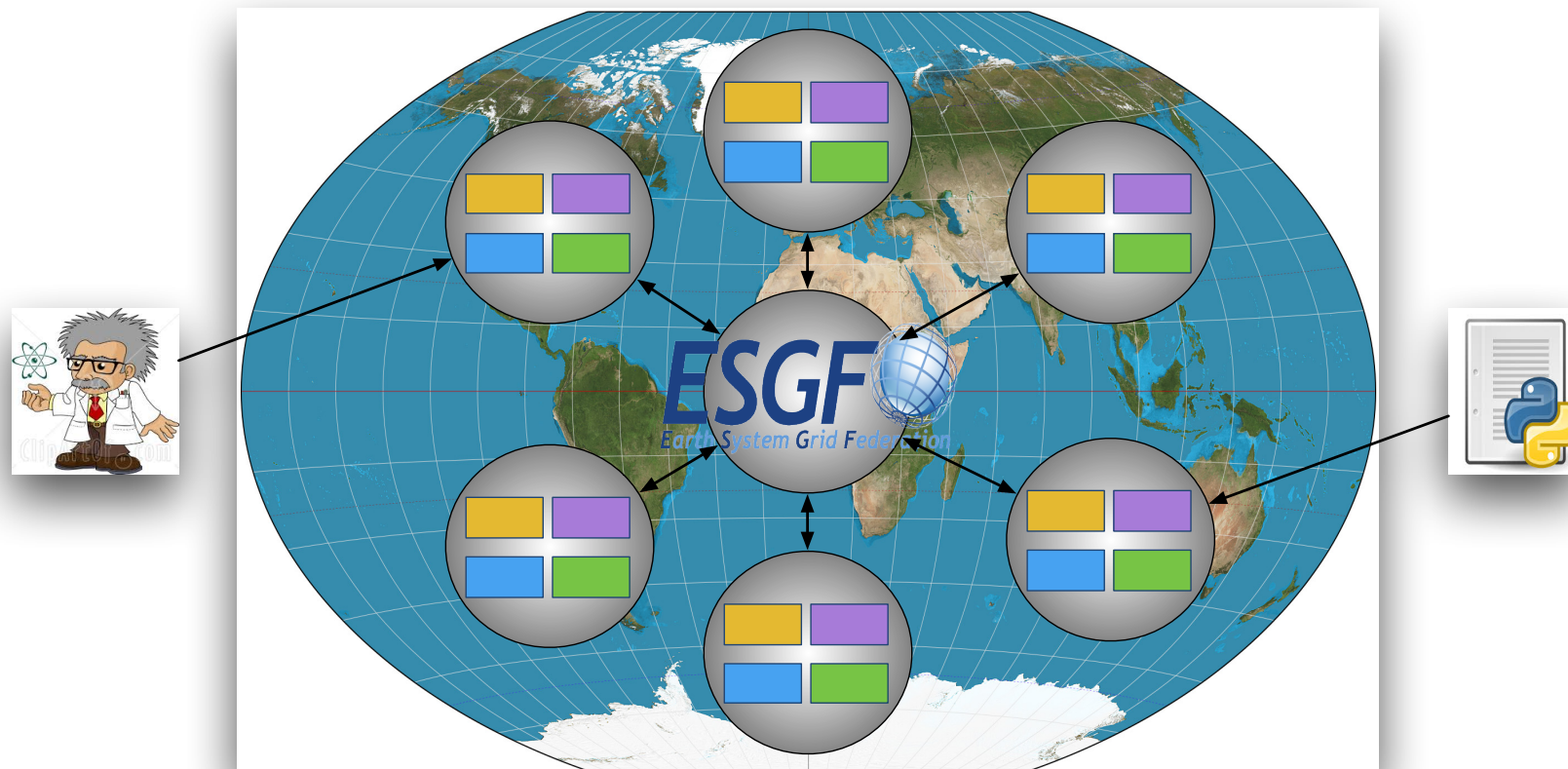
- \* Climate change is one of the most serious threats presently facing the Earth ecosystem:
  - \* New report indicates climate change is accelerating beyond earlier predictions
  - \* We are the last generation that can avoid a massive extinction of species
- \* As the world leading data infrastructure in support of climate change research, ESGF plays an important role in helping to predict and alleviate drastic climate changes



grizzly + polar = “grolar” or “pizzly”

# State of ESGF

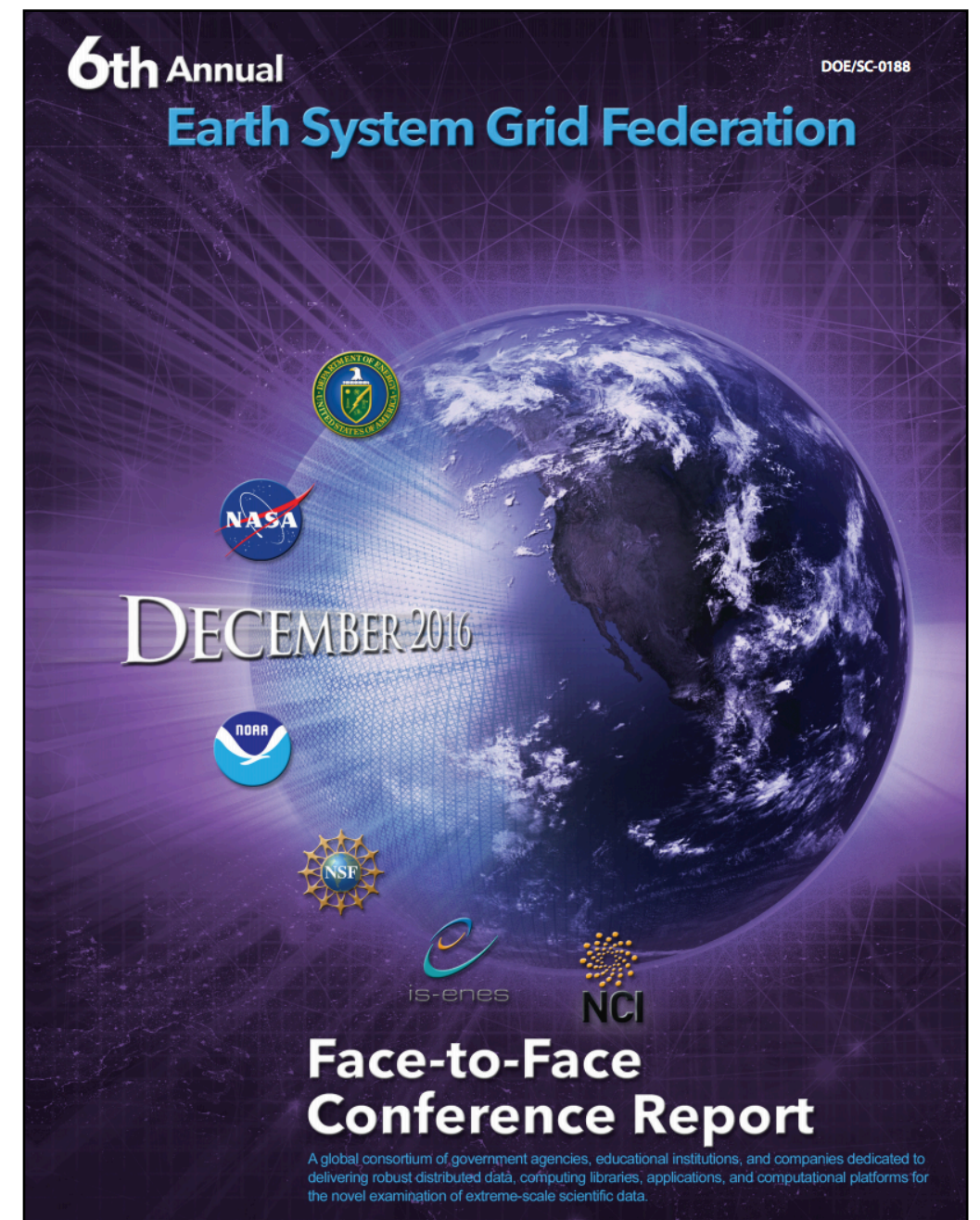
- \* ESGF has made constant, solid progress in 2018: improving the reliability of the technical infrastructure, developing new functionality, expanding data holdings and user base
- \* Most recent stats: 31 nodes, 793,026 datasets, 10,054,190 files, 133 CoG projects, 19,978 users
- \* ...but no “R&D 100” Award in 2018...





# Review of Action Items from 2017 F2F

- \* Roadmap established by ESGF-SC and ESGF-XC in 2017 for 2018 ([https://esgf.llnl.gov/esgf-media/pdf/2017-ESGF\\_F2F\\_Conference\\_Report.pdf](https://esgf.llnl.gov/esgf-media/pdf/2017-ESGF_F2F_Conference_Report.pdf)):
  - \* Short Term Plans for “CMIP6 Preparedness” (0-2 years):
    - \* Replication
    - \* Documentation and training for data publishers
    - \* Software and operations security
    - \* PID Service
    - \* Basic data reduction and analysis operations
    - \* User authentication and authorization
  - \* Longer Term Plans for ESGF longevity (2-5 years):
    - \* Server-side computation
    - \* Installation
    - \* Cloud computing
    - \* Programmatic access to data





# **2018 ACCOMPLISHMENTS**



# Preparations for CMIP6

- \* Arguably, the most important ESGF task in 2018
- \* Involved all ESGF developers, managers and node administrators, in particular:
  - \* CDNOT coordinated installation and testing of ESGF infrastructure across Nodes (S. Denvil, R. Petrie)
    - \* 5 “data challenges” held in 2018 to stress-test system with increasingly larger amounts of CMIP6 test data - very successful
    - \* Node admin guide (currently under review):
      - \* <https://docs.google.com/document/d/1y6Nd0Bea6VC6iTguFwHZr2uMM56VSpaqaXtQxCX5t-U/edit>
  - \* Replication Working Group is working at managing and improving the replication of core CMIP6 data across “Tier-1” Nodes (S. Kindermann, E. Dart)
    - \* Replica data are been published at LLNL and DKRZ with GridFTP endpoints
  - \* WIP (WGCM Infrastructure Panel) is overseeing ESGF preparations and providing connections with the CMIP modeling groups (K. Taylor, Balaji)
    - \* Recently established guidelines for configuring ESGF sites hosting CMIP6 data





# Current CMIP6 data holdings

- \* ESGF opened for CMIP6 data in June 2018
- \* Currently serving CMIP6 data from 4 Data Nodes: CNRM, GFDL, NCCS, IPSL
- \* Data replicated at LLNL, DKRZ
- \* Data holdings:
  - \* 6 CMIP6 models
  - \* ~12,835 datasets
  - \* ~37,439 files

The screenshot displays the ESGF CMIP6 Data Search interface. The browser address bar shows the URL <https://esgf-node.llnl.gov/search/cmip6/>. The page header includes the WCRP CMIP6 logo and the text "World Climate Research Programme". The user is logged in as "lucacinqui1" and is at the "ESGF@DOE/LLNL node".

The interface features a search bar with the text "Enter Text:" and a search button. Below the search bar, there are checkboxes for "Show All Replicas", "Show All Versions", and "Search Local Node Only (Including All Replicas)". The total number of results is 12835, with pagination links for "-1- 2 3 4 5 6 Next >>".

On the left side, there are several filter categories with expandable lists:

- MIP Era**: +
- Activity**: +
- Model Cohort**: +
- Product**: +
- Source ID**: -
  - ☐ CNRM-CM6-1 (3087)
  - ☐ CNRM-ESM2-1 (1963)
  - ☐ GFDL-AM4 (69)
  - ☐ GFDL-CM4 (192)
  - ☐ GISS-E2-1-G (3496)
  - ☐ IPSL-CM6A-LR (4028)
- Institution ID**: +
- Source Type**: +
- Nominal Resolution**: +
- Experiment ID**: +
- Sub-Experiment**: +
- Variant Label**: +
- Grid Label**: +
- Table ID**: +
- Frequency**: +
- Realm**: +
- Variable**: +
- CF Standard Name**: +
- Data Node**: -
  - ☐ esg1.umr-cnrm.fr (5050)
  - ☐ esgdata.gfdl.noaa.gov (261)
  - ☐ esgf.nccs.nasa.gov (3496)
  - ☐ vesg.ipsl.upmc.fr (4028)

The main content area displays a list of search results. Each result includes the dataset name, data node, version, total number of files, and full dataset services. For example, the first result is "CMIP6.CMIP.IPSL.IPSL-CM6A-LR.1pctCO2.r11p1f1.Emon.fHarvestToProduct.gr" with 1 file and services including "Show Metadata", "List Files", "THREDDS Catalog", "WGET Script", "LAS", "Show Citation", "PID", and "Globus Download".

# New ESGF Services for CMIP6

- \* ESGF has been working on providing new enhanced functionality in support of CMIP6
- \* PID (“Persistent Identifiers”) service: assigns PIDs to datasets and files at time of publication for long-term identification
  - \* Distributed service based on RabbitMQ with 3 installations at DKRZ, IPSL and LLNL
- \* Suite of web hosted services that provide an unprecedented, unmatched wealth of information about data and models that generated them
  - \* FurtherInfo URL: embedded in NetCDF files, harvested by CoG —>
  - \* ES-DOC: landing pages for datasets, models, experiments, CMIP6 —>
  - \* Errata Service: central catalog for datasets that had to be retracted for various reasons
  - \* DOI Data Citation page at WDC: provides information on how to cite the data, license, content, and related datasets (forcing). Some information is encoded as schema.org —> discoverable by
  - \* Google Dataset Search

The screenshot displays three web browser windows illustrating various ESGF services for CMIP6.

**Left Window: FurtherInfo URL**  
The browser shows the URL `https://furtherinfo.es-doc.org/static/index.html?target=CMIP6.IPSL.IPSL-CM6A-LR.1pctCO2.none.r1i1p1f1`. The page title is "CMIP6 Further Information v0.5.1.0". It features a "Support" button and a "Further Info URL" field containing the same URL. Below this, there is a section for "ES-DOC Documentation" with a table listing key information:

| Field                | Value        |
|----------------------|--------------|
| MIP Era              | CMIP6        |
| Institution          | IPSL         |
| Model                | IPSL-CM6A-LR |
| Experiment           | 1pctCO2      |
| Ensemble Description | N/A          |
| Machine Performance  | N/A          |

Below the table is a "Dataset Documentation" section with a table:

| Field               | Value   |
|---------------------|---|
| Dataset ESGF Search | N/A   |
| Dataset Errata      | N/A   |
| Dataset Citation(s) | <a href="https://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.IPSL.IPSL-CM6A-LR.1pctCO2">https://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.IPSL.IPSL-CM6A-LR.1pctCO2</a> |

At the bottom, there is an "Other Documentation" section with links to the WCRP CMIP6 Homepage and the ES-DOC CMIP6 Homepage.

**Middle Window: Dataset Errata - Search**  
The browser shows the URL `https://errata.es-doc.org/static/index.html`. The page title is "Dataset Errata - Search v0.6.1.0". It features a "Support" button, a "PID" button, and a "Login" button. Below these buttons is a search form with fields for Project, Experiment ID, Institution ID, Source ID, Variable ID, Severity, and Status. Below the search form is a table listing 16 issues:

| #  | Institute    | Title   | Created    | Updated    | Closed | Severity | Status   |
|----|--------------|---|------------|------------|--------|----------|----------|
| 1  | CNRM-CERFACS | Wrong realm ccnBgChem typo                                | 2018-11-14 | 2018-11-16 | --     | Low      | Resolved |
| 2  | NOAA-GFDL    | Incorrect some coordinates and cell_methods in piCont ... | 2018-11-08 | 2018-11-08 | --     | Medium   | New      |
| 3  | NOAA-GFDL    | Error in variable "comment" metadata                      | 2018-11-01 | 2018-11-16 | --     | Low      | New      |
| 4  | NOAA-GFDL    | alibscop erroneous data units                             | 2018-10-29 | 2018-11-16 | --     | Low      | New      |
| 5  | IPSL         | 300 years extension for abrupt-4xCO2                      | 2018-10-22 | 2018-10-22 | --     | Low      | Resolved |
| 6  | IPSL         | Irrelevant CFC in experiment other than historical        | 2018-10-19 | 2018-10-23 | --     | Low      | Resolved |
| 7  | IPSL         | Instabilities which lead to erroneous values of tas a ... | 2018-10-16 | 2018-10-16 | --     | Critical | On Hold  |
| 8  | IPSL         | tas instabilities lead to erroneous values of tasmax      | 2018-10-05 | 2018-10-16 | --     | Critical | On Hold  |
| 9  | IPSL         | Versioning errors for 1pctCO2 and abrupt-4xCO2            | 2018-07-27 | 2018-07-27 | --     | Critical | Resolved |
| 10 | IPSL         | Wrong realm "ccnBgChem" typo                              | 2018-07-26 | 2018-08-08 | --     | Low      | Resolved |
| 11 | IPSL         | Unchanged PIDs for new version                            | 2018-07-20 | 2018-07-21 | --     | High     | Resolved |
| 12 | IPSL         | Some sea ice variables in 3D instead of 1D                | 2018-07-12 | 2018-07-17 | --     | Low      | Resolved |
| 13 | IPSL         | Integers instead of PFTs names                            | 2018-07-02 | 2018-10-12 | --     | Low      | Resolved |
| 14 | IPSL         | Integers instead of ocean passages names                  | 2018-07-02 | 2018-07-17 | --     | Low      | Resolved |
| 15 | IPSL         | "area.coordinates" attribute is missing                   | 2018-07-02 | 2018-07-17 | --     | Low      | Resolved |
| 16 | IPSL         | Time instantaneous data with time boundaries              | 2018-07-02 | 2018-07-13 | --     | Low      | Wont Fix |

**Right Window: Metadata for 'CMIP6.CMIP.IPSL.IPSL-CM6A-LR.1pctCO2'**  
The browser shows the URL `https://cera-www.dkrz.de/WDCC/ui/cersearch/cmip6?input=CMIP6.CMIP.IPSL.IPSL-CM6A-LR.1pctCO2`. The page title is "Metadata for 'CMIP6.CMIP.IPSL.IPSL-CM6A-LR.1pctCO2'". It features a "Support" button, a "PID" button, and a "Login" button. Below these buttons is a section for "General Information" with a table listing key information:

| Field          | Value  |
|----------------|--|
| Name           | CMIP6.CMIP.IPSL.IPSL-CM6A-LR.1pctCO2   |
| Abstract       | Coupled Model Intercomparison Project Phase 6 (CMIP6) data sets. These data includes all datasets published for 'CMIP6.CMIP.IPSL.IPSL-CM6A-LR.1pctCO2' according to the Data Reference Syntax defined as 'mip_era_activity_id_institution_id_source_id_experiment_id_member_id_table_id_variable_id_grid_label_version'. |
| Subjects       | CMIP6 climate  |
| Rights License | Creative Commons Attribution 4.0 International License (CC BY-SA 4.0)  |



# New ESGF Services for CMIP6

- \* ESGF has been working on providing new enhanced functionality in support of CMIP6
- \* PID (“Persistent Identifiers”) service: assigns PIDs to datasets and files at time of publication for long-term identification
  - \* Distributed service based on RabbitMQ with 3 installations at DKRZ, IPSL and LLNL
- \* ES-DOC: eco-system of services that provide detailed documentation for models, experiments
  - \* Hyperlinked by FurtherInfo URL
  - \* Includes Errata Service - central catalog for datasets that had to be retracted for various reasons
- \* Citation and DOI service: provides additional information on the model, data and how to cite
- \* Overall, unprecedented, unmatched wealth of information about data and models that generated them

**Further Info URL:** <https://furtherinfo.es-doc.org/CMIP6.IPSL.IPSL-CM6A-LR.1pctCO2.none.r1i1p1f1>

**ES-DOC Documentation**

|                      |              |
|----------------------|--------------|
| MIP Era              | CMIP6        |
| Institution          | IPSL         |
| Model                | IPSL-CM6A-LR |
| Experiment           | 1pctCO2      |
| Ensemble Description | N/A          |
| Machine Performance  | N/A          |

**Dataset Documentation**

|                     |   |
|---------------------|---|
| Dataset ESGF Search | N/A   |
| Dataset Errata      | N/A   |
| Dataset Citation(s) | <a href="https://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.IPSL.IPSL-CM6A-LR.1pctCO2">https://cera-www.dkrz.de/WDCC/meta/CMIP6/CMIP6.CMIP.IPSL.IPSL-CM6A-LR.1pctCO2</a> |

**Other Documentation**

|                       |   |
|-----------------------|---|
| WCRP CMIP6 Homepage   | <a href="https://www.wcrp-climate.org/wgcm-cmip/wgcm-cmip6">https://www.wcrp-climate.org/wgcm-cmip/wgcm-cmip6</a> |
| ES-DOC CMIP6 Homepage | <a href="https://es-doc.org/cmip6">https://es-doc.org/cmip6</a>   |

**Dataset Errata - Search** v0.6.1.0

| Project | Experiment ID | Institution ID | Source ID | Variable ID | Severity | Status |
|---------|---------------|----------------|-----------|-------------|----------|--------|
| CMIP6   | *             | *              | *         | *           | *        | *      |

Total Issues = 16. Filtered Issues = 16.

| #  | Institute    | Title   | Created    | Updated    | Closed | Severity | Status   |
|----|--------------|---|------------|------------|--------|----------|----------|
| 1  | CNRM-CERFACS | Wrong realm ccnBgChem typo                                | 2018-11-14 | 2018-11-16 | --     | Low      | Resolved |
| 2  | NOAA-GFDL    | Incorrect some coordinates and cell_methods in piCont ... | 2018-11-08 | 2018-11-08 | --     | Medium   | New      |
| 3  | NOAA-GFDL    | Error in variable "comment" metadata                      | 2018-11-01 | 2018-11-16 | --     | Low      | New      |
| 4  | NOAA-GFDL    | albiscop erroneous data units                             | 2018-10-29 | 2018-11-16 | --     | Low      | New      |
| 5  | IPSL         | 300 years extension for abrupt-4xCO2                      | 2018-10-22 | 2018-10-22 | --     | Low      | Resolved |
| 6  | IPSL         | Irrelevant CFC in experiment other than historical        | 2018-10-19 | 2018-10-23 | --     | Low      | Resolved |
| 7  | IPSL         | Instabilities which lead to erroneous values of tas ...   | 2018-10-16 | 2018-10-16 | --     | Critical | On Hold  |
| 8  | IPSL         | tas instabilities lead to erroneous values of tasmax      | 2018-10-05 | 2018-10-16 | --     | Critical | On Hold  |
| 9  | IPSL         | Versioning errors for 1pctCO2 and abrupt-4xCO2            | 2018-07-27 | 2018-07-27 | --     | Critical | Resolved |
| 10 | IPSL         | Wrong realm "ccnBgChem" typo                              | 2018-07-26 | 2018-08-08 | --     | Low      | Resolved |
| 11 | IPSL         | Unchanged PIDs for new version                            | 2018-07-20 | 2018-07-21 | --     | High     | Resolved |
| 12 | IPSL         | Some sea ice variables in 3D instead of 1D                | 2018-07-12 | 2018-07-17 | --     | Low      | Resolved |
| 13 | IPSL         | Integers instead of PFTs names                            | 2018-07-02 | 2018-10-12 | --     | Low      | Resolved |
| 14 | IPSL         | Integers instead of ocean passages names                  | 2018-07-02 | 2018-07-17 | --     | Low      | Resolved |
| 15 | IPSL         | "area.coordinates" attribute is missing                   | 2018-07-02 | 2018-07-17 | --     | Low      | Resolved |
| 16 | IPSL         | Time instantaneous data with time boundaries              | 2018-07-02 | 2018-07-13 | --     | Low      | Wont Fix |

Total Issues = 16. Filtered Issues = 16.

**Metadata for 'CMIP6.CMIP.IPSL.IPSL-CM6A-LR.1pctCO2'**

**General Information**

**Name** CMIP6.CMIP.IPSL.IPSL-CM6A-LR.1pctCO2

**Abstract** Coupled Model Intercomparison Project Phase 6 (CMIP6) data sets. These data includes all datasets published for 'CMIP6.CMIP.IPSL.IPSL-CM6A-LR.1pctCO2' according to the Data Reference Syntax defined as 'mip\_era\_activity\_id\_institution\_id\_source\_id\_experiment\_id\_member\_id\_table\_id\_variable\_id\_grid\_label\_version'.

The Earth System Model (IPSL-CM6A-LR, released in 2017, includes the components: atmos: LM3Z (NPv6, N96, 144 x 143 longitude/latitude; 79 levels; top level 40000 m), land: ORCHIDEE (v2.0, Water/Carbon/Energy mode), ocean: NEMO-OPA (eORCA1.3, tripolar primarily 1deg; 362 x 332 longitude/latitude; 75 levels; top grid cell 0-2 m), ccnBgchem: NEMO-PISCES, seaIce: NEMO-LIM3.

The model was run by the Institut Pierre Simon Laplace, Paris 75252, France (IPSL) in native nominal resolutions: atmos: 250 km, land: 250 km, ocean: 100 km, ccnBgchem: 100 km, seaIce: 100 km.

Project: These data have been generated as part of the internationally-coordinated Coupled Model Intercomparison Project Phase 6 (CMIP6); see also GMD Special Issue: [http://www.gosci-model-dev.net/special\\_issue590.html](http://www.gosci-model-dev.net/special_issue590.html). The simulation data provides a basis for climate research designed to answer fundamental science questions, and the results will undoubtedly be relied on by authors of the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC-AR6).

CMIP6 is a project coordinated by the Working Group on Coupled Modelling (WGCM) as part of the World Climate Research Programme (WCRP). Phase 6 builds on previous phases executed under the leadership of the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and relies on the Earth System Grid Federation (ESGF) and the Centre for Environmental Data Analysis (CEDA) along with numerous related activities for implementation. The original data is hosted and partially replicated at a federated collection of data nodes, and most of the data relied on by the IPCC is being archived for long-term preservation at the IPCC Data Distribution Centre (IPCC DDC) hosted by World Data Centre for Climate (WDCC) at DKRZ.

The project includes simulations from about 90 global climate models and around 40 institutions and organizations worldwide. - Project website: <https://pcmdi.lnl.gov/CMIP6>

**Subjects** CMIP6, climate

**Rights License** Creative Commons Attribution 4.0 International License (CC BY-SA 4.0)

CMIP6 model data is evolving in the sense that datasets are changed and added as new versions. The author list and the title are not final, either. Cite this data collection including the latest dataset version according to the Data Citation Guidelines (<http://bit.ly/2gBCuqH>). Individuals using the data must abide to the terms of use for CMIP6 data (<https://pcmdi.lnl.gov/CMIP6/TermsOfUse>). Details on any license restrictions are recorded as global attributes in the files.

# Obs4MIPs Update

- \* ESGF is also increasing availability and support for observations data through its Obs4MIPs program (R. Ferraro, P. Gleckler, D. Waliser and P. Durack)
- \* Obs4MIPs are selected observational datasets from NASA, ESA, NOAA, etc. that follow the same data and metadata conventions as CMIP model output, intended for easier comparison and validation of model data
- \* Progress in 2018:
  - \* Additional datasets from several agencies (90 datasets total)
  - \* Introduced support for “dataset indicators” that represent the “maturity level” for model evaluation
  - \* New specification for dir structure, filenames, search facets that is aligned with CMIP6 (ODSv2.1)
  - \* Datasets are being moved and republished at LLNL to follow the new specs
  - \* Obs4MIPs CoG site moved to LLNL for long-term support

| Technical Requirements  |  | Dataset Suitability and Maturity  |  |  | Comparison Complexity   |
|---|--|---|--|--|---|
| Meets obs4MIPs data technical requirements                                  | Includes obs4MIPs technical note information                   | Closeness or robustness of measurement to observed reference quantity   | Maturity with respect to climate model evaluation  | Provision for robust uncertainty information   | Complexity of Model Observation Comparison  |
| Data suitably processed with CMOR and/or consistent with obs4MIPs standards | Complete technical note information provided                   | Firmly established and/or validated methodology   | Multiple peer-reviewed examples of application to CMIP climate model evaluation                              | Uncertainty information provided per retrieval/grid point  | Comparison can be made directly with CMIP model output variable   |
| Largely complete with minor metadata inconsistencies                        | Technical note information incomplete and/or could be improved | Indirect means of calculation or observations only providing partial constraint (e.g. ocean surface latent heat flux) | One peer-reviewed example of application to CMIP climate and/or examples of other sorts of model evaluation. | General uncertainty information given relative to the methodology and dataset as a whole - backed by actual field/in-situ validation exercises | Comparison requires some simple post processing of CMIP output variable(s) (e.g. vertical integral or ratio of two variables) |
| Non-compliant. Should be removed from database!                             | Technical note not provided                                    | Largely model-derived quantity (e.g. LAI, root zone soil moisture, NPP)   | As of DATE-TBS, no significant application to climate model evaluation                                       | No uncertainty information provided  | Comparison requires complex processing of CMIP output (e.g. "simulator", budget calculation)                                  |

Navigation bar: Nodes | Pages | Content Indices | x obs4MIPs Data | x cmip6 Data S... | x ES-DOC - Fur... | x unprecident... | x ES-DOC - V... | x https://esgf-node.lnl.gov/search/obs4mips/?template=obs4mips&limit=200

Hosted by Department of Energy Lawrence Livermore National Laboratory

Powered by ESGF and Logos Welcome, lucacine1! | My Profile | Log out

Home About Us Governance Contact Us

### Obs4MIPs Data Table

The following table is automatically updated with Obs4MIPs datasets available throughout the Earth System Grid Federation.

| Instrument         | Dataset Name                                    | Variables  | Tech Note | Maturity Indicators |
|--------------------|---|--|-----------|---------------------|
| AIRS               | obs4mips.NASA-JPLAIRS.hus.mon                   | Specific Humidity<br>Specific Humidity Number of Observations<br>Specific Humidity Standard Error  |           | [■ ■ ■ ■ ■]         |
| AIRS               | obs4mips.NASA-JPLAIRS.ta.mon                    | Air Temperature<br>Air Temperature Number of Observations<br>Air Temperature Standard Error  |           | [■ ■ ■ ■ ■]         |
| AIRS-TA-2.0        | obs4MIPs.NASA-JPLAIRS-TA-2.0.mon.hurStdErr.gn   | Air Temperature number of observations   |           |                     |
| AIRS-TA-2.0        | obs4MIPs.NASA-JPLAIRS-TA-2.0.mon.hurNobs.gn     | Relative humidity standard error   |           |                     |
| AIRS-TA-2.0        | obs4MIPs.NASA-JPLAIRS-TA-2.0.mon.hurStdErr.gn   | Relative Humidity number of observations   |           |                     |
| AIRS-TA-2.0        | obs4MIPs.NASA-JPLAIRS-TA-2.0.mon.hus.gn         | Specific Humidity  |           |                     |
| AIRS-TA-2.0        | obs4MIPs.NASA-JPLAIRS-TA-2.0.mon.hur.gn         | Relative Humidity  |           |                     |
| AIRS-TA-2.0        | obs4MIPs.NASA-JPLAIRS-TA-2.0.mon.taStdErr.gn    | Air Temperature standard error   |           |                     |
| AIRS-TA-2.0        | obs4MIPs.NASA-JPLAIRS-TA-2.0.mon.ta.gn          | Air Temperature  |           |                     |
| AIRS-TA-2.0        | obs4MIPs.NASA-JPLAIRS-TA-2.0.mon.taNobs.gn      | Air Temperature number of observations   |           |                     |
| AMSR               | obs4mips.REMSR.AMSR.tas.mon                     | Sea Surface Temperature<br>Sea Surface Temperature Number of Observations<br>Sea Surface Temperature Standard Error  |           | [■ ■ ■ ■ ■]         |
| ARC-SST-1.1        | obs4MIPs.UOE.ARC-SST-1.1.mon                    | sea surface temperature  |           |                     |
| ATSR2-AATSR        | obs4mips.SU.ATSR2-AATSR.od55oaser.mon           | latitude<br>longitude<br>Ambient Aerosol Optical Thickness at 550 nm<br>time   |           |                     |
| AVISO              | obs4mips.CNES.AVISO.zos.mon                     | Sea Surface Height Above Geoid<br>Sea Surface Height Above Geoid Number of Observations<br>Sea Surface Height Above Geoid Standard Error   |           | [■ ■ ■ ■ ■]         |
| CERES-EBAF         | obs4MIPs.NASA-LaRC.CERES-EBAF.atmos.mon         | TOA Outgoing Longwave Radiation<br>TOA Outgoing Clear-Sky Longwave Radiation<br>TOA Incident Shortwave Radiation<br>TOA Outgoing Shortwave Radiation<br>TOA Outgoing Clear-Sky Shortwave Radiation                                 |           |                     |
| CERES-EBAF_Surface | obs4MIPs.NASA-LaRC.CERES-EBAF_Surface.atmos.mon | Surface Downwelling Longwave Radiation<br>Surface Downwelling Clear-Sky Longwave Radiation<br>Surface Upwelling Longwave Radiation<br>Surface Downwelling Shortwave Radiation<br>Surface Downwelling Clear-Sky Shortwave Radiation |           |                     |



# Installation of ESGF Node

- \* ESGF has worked at improving the process of installing, maintaining and securing an ESGF Node, in several directions:
  - \* “Classic” shell-based installer was upgraded to support 6 major ESGF releases in 2018: 2.6.5, 2.6.7, 2.6.8b, 2.6.9, 2.7.1, 2.8 (S. Ames, P. Dwarakanath)
    - \* New “esgf-scanner” tool that automatically creates a manifest of all software packages included in an ESGF release and lists all CVEs
    - \* Software stack now working on Centos7
  - \* Upcoming migration to new Python-based installer (3.0) (W. Hill) - now in beta
    - \* Will be much easier to run, test and maintain than current shell installer
  - \* New Continuous Integration (CI) system built on Jenkins provides automatic builds and runs the test-suite for every GitHub commit (S. Gardoll)

**Jenkins**

Search [ ] Log In

ENABLE AUTO REFRESH

| All              | devrel     | esgf-docker                           | master             | pipeline-devrel      | pipeline-master |
|------------------|------------|---------------------------------------|--------------------|----------------------|-----------------|
| S                | W          | Name ↓                                | Last Success       | Last Failure         | Last Duration   |
| [Folder Icon]    | [Sun Icon] | CVEChecker                            | N/A                | N/A                  | N/A             |
| [Blue Ball Icon] | [Sun Icon] | esgf-crm_devel                        | 10 hr - #2651      | 10 days - #2631      | 24 sec          |
| [Blue Ball Icon] | [Sun Icon] | esgf-crm_master                       | 4 mo 27 days - #19 | N/A                  | 25 sec          |
| [Blue Ball Icon] | [Sun Icon] | esgf-search_devel                     | 10 hr - #2803      | 17 days - #2769      | 22 sec          |
| [Blue Ball Icon] | [Sun Icon] | esgf-search_master                    | 4 mo 27 days - #19 | N/A                  | 25 sec          |
| [Blue Ball Icon] | [Sun Icon] | esgf-dashboard_devel                  | 10 hr - #2639      | 3 mo 13 days - #2433 | 13 sec          |
| [Blue Ball Icon] | [Sun Icon] | esgf-dashboard_master                 | 4 mo 27 days - #18 | N/A                  | 15 sec          |
| [Grey Ball Icon] | [Sun Icon] | esgf-desktop_devel                    | 1 yr 0 mo - #2209  | 1 yr 2 mo - #2111    | 11 sec          |
| [Grey Ball Icon] | [Sun Icon] | esgf-desktop_master                   | 2 yr 5 mo - #14    | N/A                  | 21 sec          |
| [Blue Ball Icon] | [Sun Icon] | esgf-getcert_devel                    | 10 hr - #2594      | 3 mo 13 days - #2388 | 4.7 sec         |
| [Blue Ball Icon] | [Sun Icon] | esgf-getcert_master                   | 4 mo 27 days - #16 | N/A                  | 5.2 sec         |
| [Blue Ball Icon] | [Sun Icon] | esgf-sls_devel                        | 10 hr - #2656      | 10 days - #2636      | 19 sec          |
| [Blue Ball Icon] | [Sun Icon] | esgf-sls_master                       | 4 mo 27 days - #18 | N/A                  | 22 sec          |
| [Blue Ball Icon] | [Sun Icon] | esgf-node-manager_devel               | 10 hr - #2600      | 17 days - #2566      | 26 sec          |
| [Blue Ball Icon] | [Sun Icon] | esgf-node-manager_master              | 4 mo 27 days - #19 | N/A                  | 32 sec          |
| [Blue Ball Icon] | [Sun Icon] | esgf-security_devel                   | 10 hr - #2625      | 17 days - #2591      | 24 sec          |
| [Blue Ball Icon] | [Sun Icon] | esgf-security_master                  | 4 mo 27 days - #17 | N/A                  | 1 min 9 sec     |
| [Blue Ball Icon] | [Sun Icon] | esgf_docker_multibranch_pipeline      | 39 sec - log       | N/A                  | 7.3 sec         |
| [Blue Ball Icon] | [Sun Icon] | test-esgf-docker_multibranch_pipeline | 4 mo 4 days - log  | N/A                  | 3.5 sec         |

icon: S M L

Legend RSS for all RSS for failures RSS for just latest builds

The screenshot displays the Jenkins Pipeline Builder interface. At the top, the browser address bar shows the URL: `https://esgf-build.jpel.upmc.fr/jenkins/view/pipeline-master/`. The Jenkins logo and navigation tabs are visible. The main heading is "Build Pipeline".

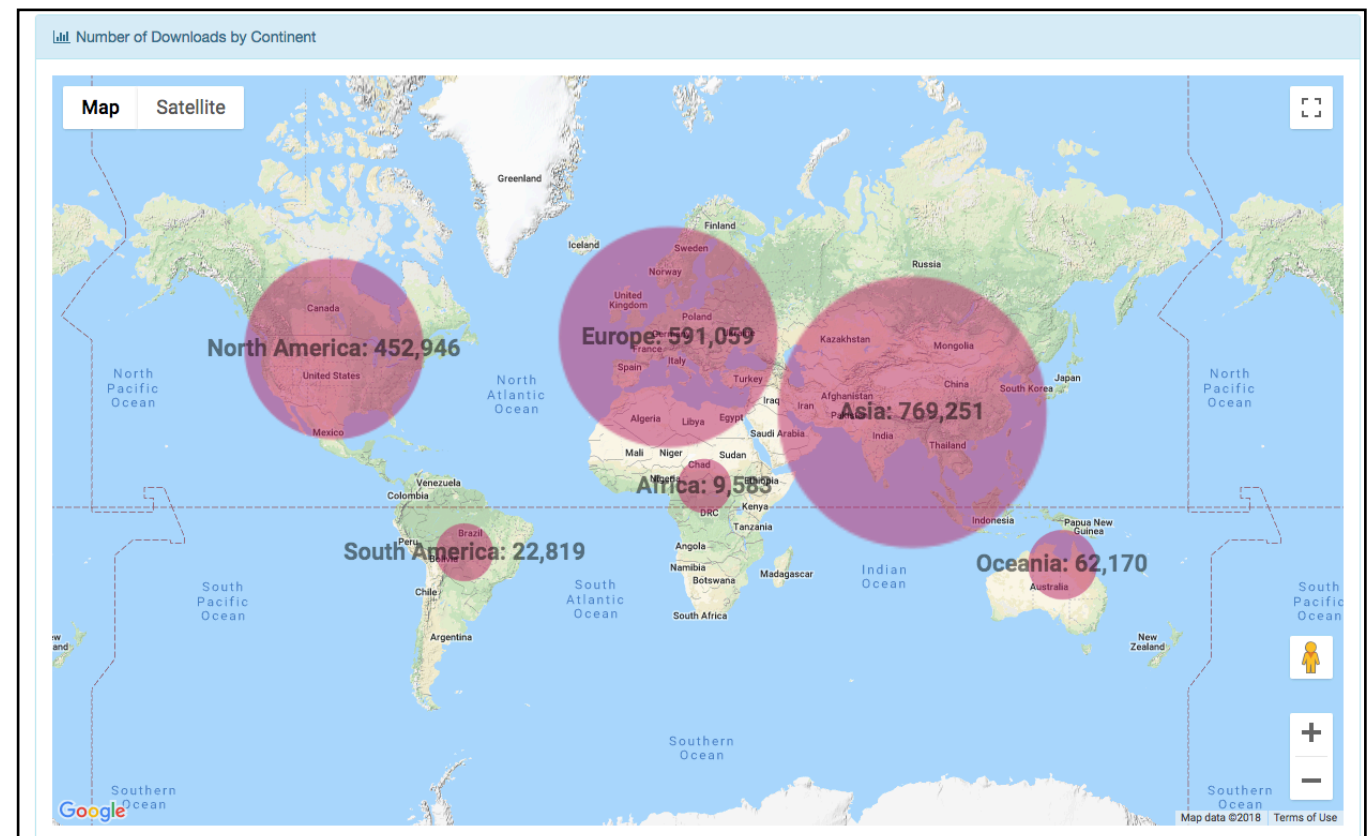
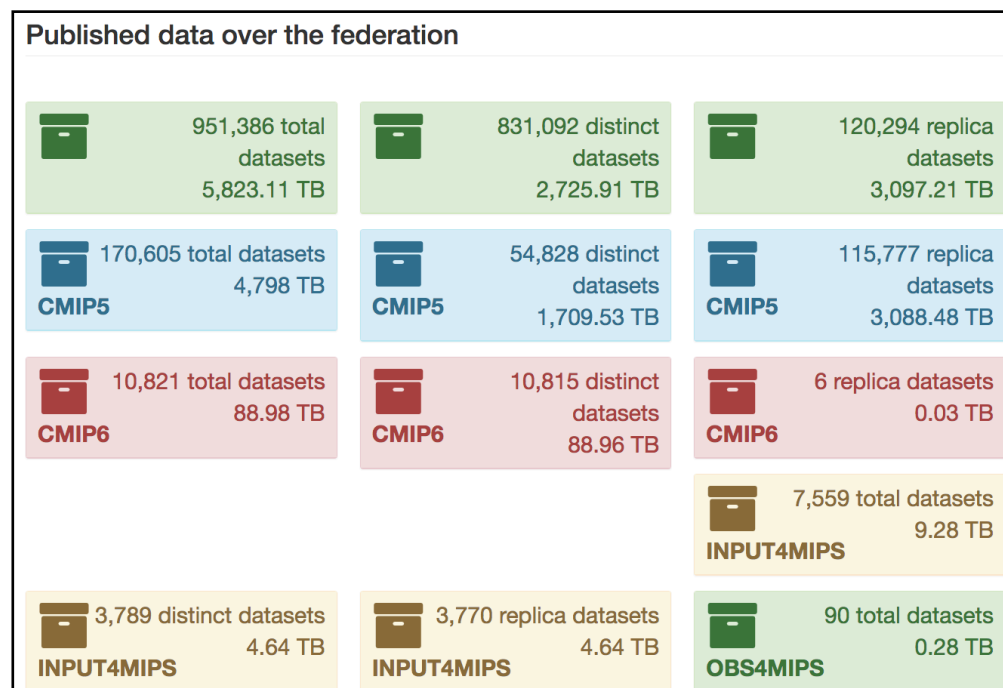
The pipeline diagram is set against an orange background and shows the following stages and tasks:

- Stage #19: esgf-node-manager\_master** (Green box)
  - Task 1: 11 jun 2018 14:05:10, 32 sec, esgf (green box)
- Stage #18: esgf-dashboard\_master** (Green box)
  - Task 1: 11 jun 2018 14:05:05, 16 sec, esgf (green box)
- Stage esgf-desktop\_master** (Orange box)
  - Task 1: 11 jun 2018 14:05:05, 16 sec, esgf (green box)
- Stage #16: esgf-getcert\_master** (Green box)
  - Task 1: 11 jun 2018 14:05:05, 8.3 sec, esgf (green box)
- Stage #19: esgf-security\_master** (Green box)
  - Task 1: 11 jun 2018 14:05:05, 3 min 8 sec, esgf (green box)
- Stage #19: esgf-orp\_master** (Green box)
  - Task 1: 11 jun 2018 14:07:05, 25 sec, esgf (green box)
- Stage #18: esgf-scp\_master** (Green box)
  - Task 1: 11 jun 2018 14:07:05, 50 sec, esgf (green box)
- Stage #19: esgf-search\_master** (Green box)
  - Task 1: 11 jun 2018 14:07:40, 25 sec, esgf (green box)

Arrows indicate the flow between stages. The bottom section shows a partial view of another pipeline with stages #16, #17, and #18.

# Other ESGF Development

- \* Idea Working Team has been progressing on transitioning the current ESGF Security infrastructure (based on OpenID 2.0) to more current industry standards: OAuth and OpenID-Connect (P. Kershaw)
  - \* New software components SLCS, ESGF-Auth to replace current IdP, ORP, MyProxy
- \* ESGF publisher and ESG prep - several upgrades to support CMIP6 and improve performance (S. Ames, G. Levavasseur)
- \* Dashboard team - integrating the information provider into the installer and supporting the central metrics aggregator site, also developing custom view for CMIP6 (S. Fiore, A. Nuzzo, M. Mirto)
- \* ...and much more...



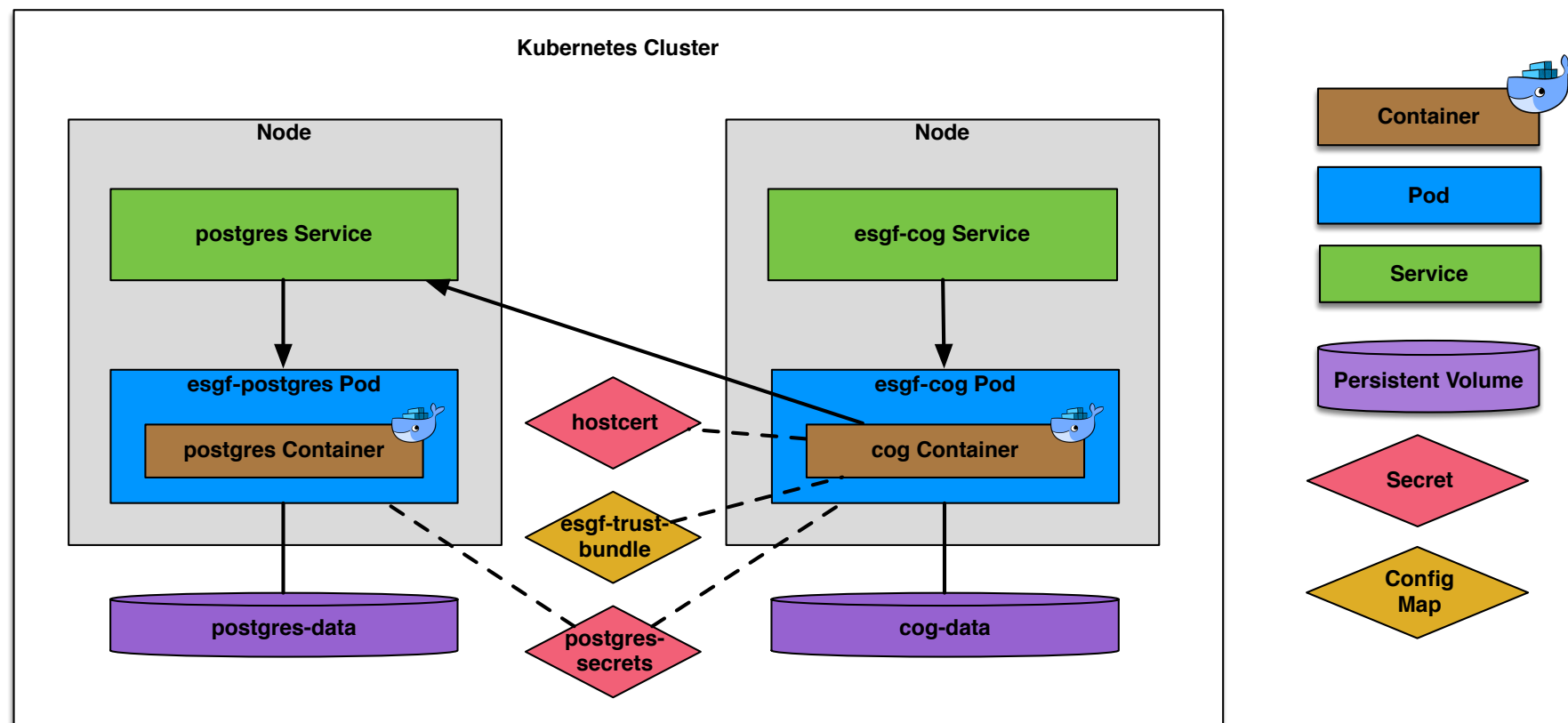


# ESGF NEW DIRECTIONS



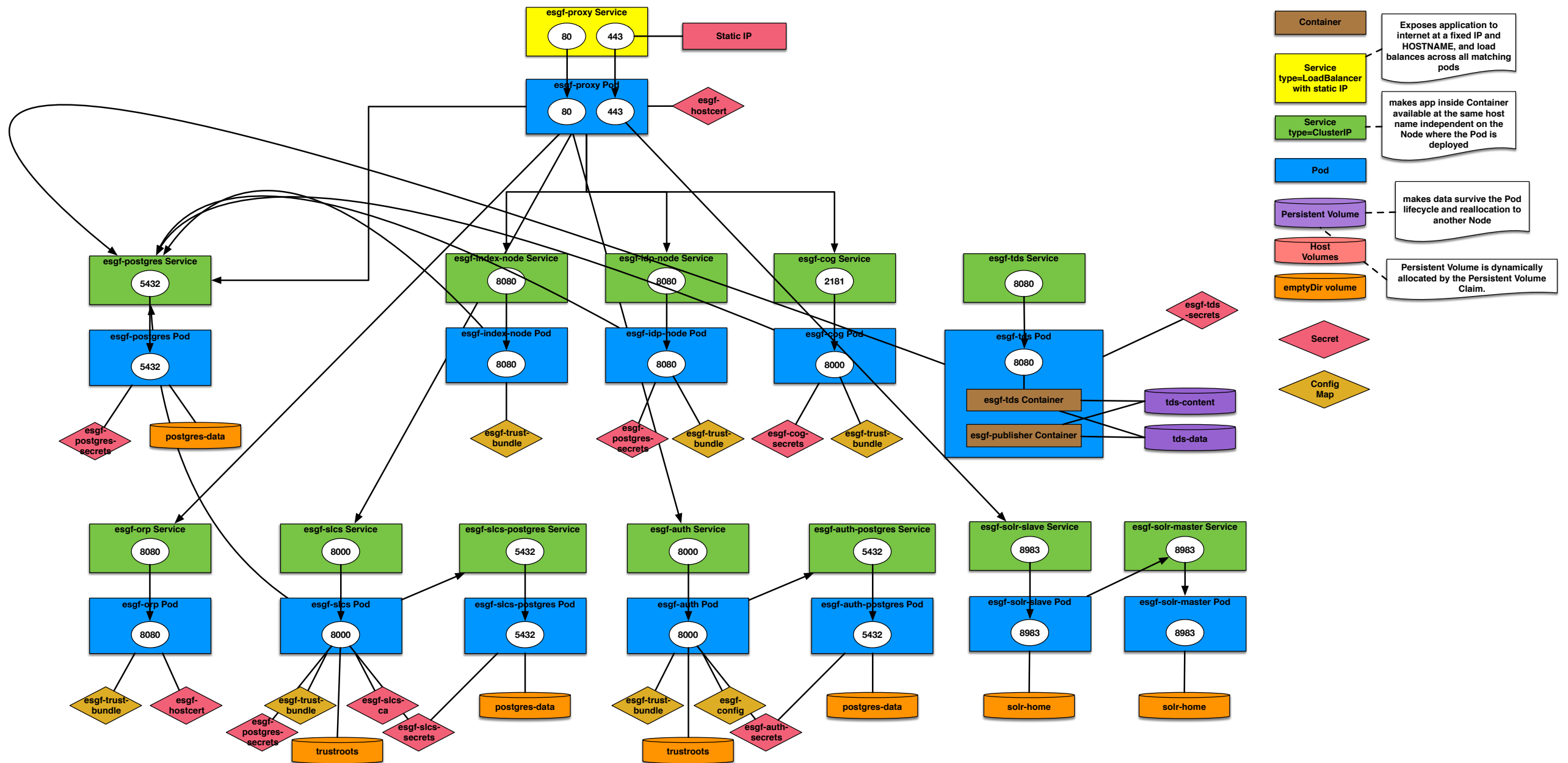
# Containerization

- \* ESGF/Docker: alternative architecture for ESGF Node where all services are packaged, deployed and managed as Docker containers
- \* Advantages of container based architecture (“micro-services”):
  - \* Easier to deploy and test
  - \* More flexible
  - \* More scalable
  - \* Easier to evolve
- \* ESGF/Docker first release in September 2018
  - \* Stable but not feature complete (no Globus)
  - \* Based on Docker, Kubernetes and Helm



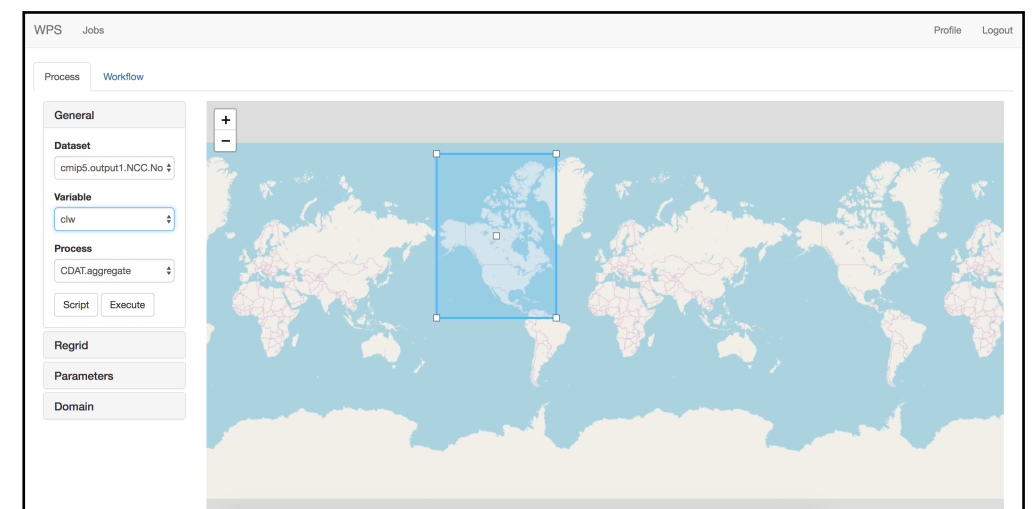
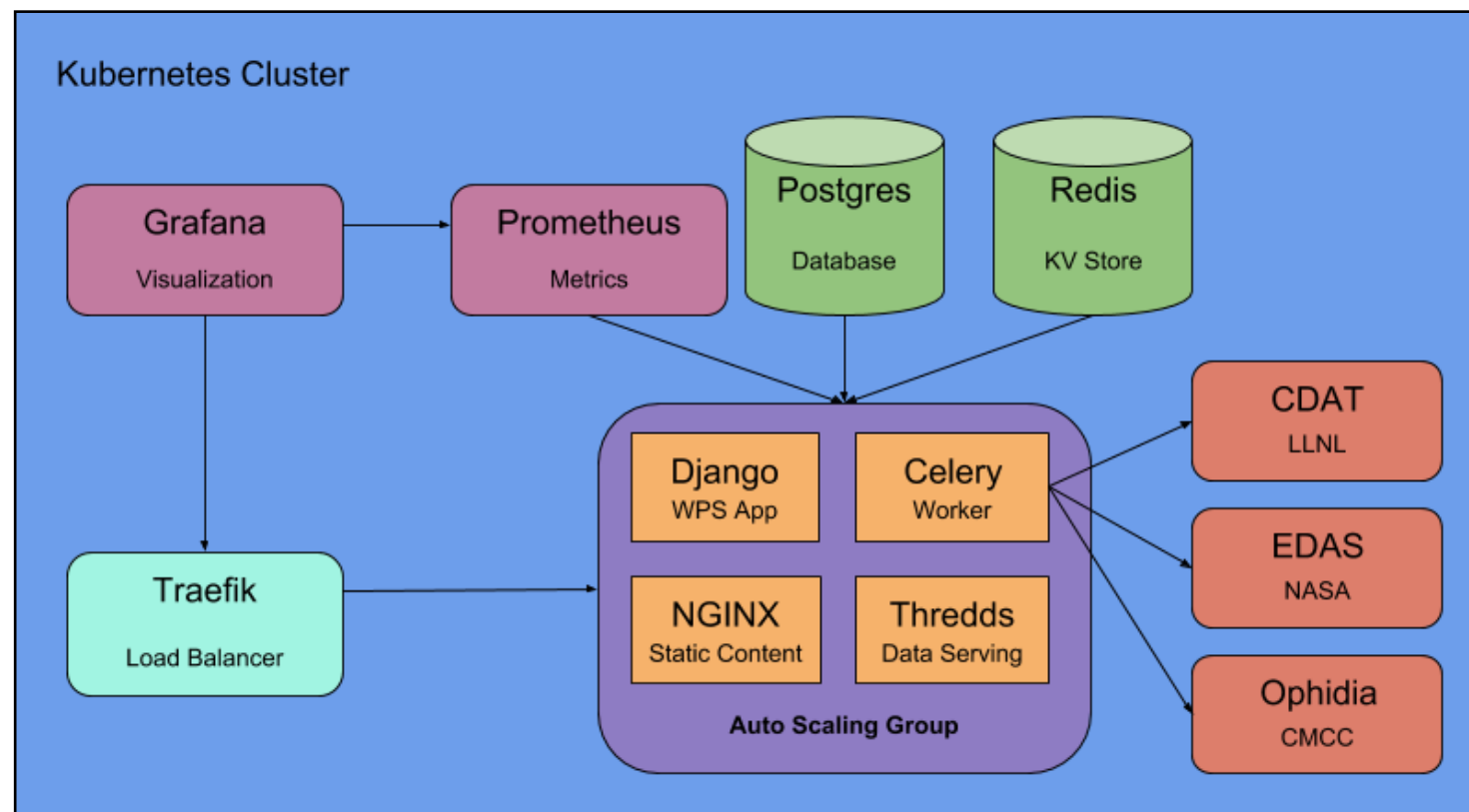


# ESGF/Docker Complete Software Stack



# Compute Node

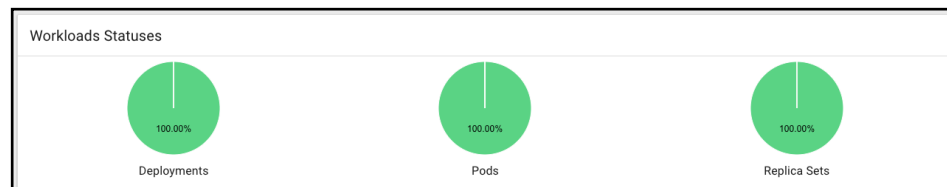
- \* The ESGF Compute Working Team has made great progress in developing scalable computing capabilities for ESGF (J. Boutte, C. Doutriaux, T. Maxwell)
- \* Architecture of compute node was designed from the ground up as a system of interacting Docker containers
  - \* Highly scalable - both horizontally and vertically
- \* 3 alternative back-ends implement the same algorithms (sub-set, average, min/max, etc.)
  - \* CDAT, Ophidia, EDAS
- \* Status: already converted to Kubernetes+Helm, ready to be deployed alongside ESG/Docker stack





# Moving to the Cloud

- \* ESGF is experimenting with moving its services -all or in part- to the Cloud
- \* Cloud advantages: practically unlimited scalability, high availability, managed resources
- \* Cloud challenges: cost model, new architecture designs
  - \* How to persist data, how to plan for failure, how to run on a cluster of nodes
- \* Several efforts undergoing:
  - \* ESGF/Docker with Kubernetes is immediately suitable for Cloud deployment
  - \* GFDL is running a prototype node on Google GKE, published some CMIP6 data, enabling access to Pangeo via openDAP
  - \* GSFC/JPL planning to deploy a single ESGF/NASA node on AWS GovCloud
  - \* New Index Node architecture based on Solr Cloud, stable deployment on ASS for several months



| Name                               | Node  | Status  | Restarts | Age        |
|------------------------------------|---|---------|----------|------------|
| esgf-index-node-864c6448c7-c6fk    | ip-192-168-195-203.us-west-2.compute.internal | Running | 0        | 21 minutes |
| esgf-slcs-85b9ff9d4-fhkvh          | ip-192-168-195-203.us-west-2.compute.internal | Running | 0        | 33 minutes |
| esgf-postgres-slcs-57b5c97f47-4hmx | ip-192-168-195-203.us-west-2.compute.internal | Running | 0        | 33 minutes |
| esgf-ids-756845bb9-48p8f           | ip-192-168-157-180.us-west-2.compute.internal | Running | 0        | 2 hours    |
| esgf-orp-6fc779df47-z5z7r          | ip-192-168-195-203.us-west-2.compute.internal | Running | 0        | 3 hours    |
| esgf-cog-67486c98d5-smdlc          | ip-192-168-157-180.us-west-2.compute.internal | Running | 0        | 3 hours    |
| esgf-postgres-cog-78ff645bf6-5m2zm | ip-192-168-195-203.us-west-2.compute.internal | Running | 0        | 3 hours    |
| esgf-ldp-node-556b5d567b-18s7q     | ip-192-168-157-180.us-west-2.compute.internal | Running | 0        | 3 hours    |
| esgf-proxy-784589854d-79hdj        | ip-192-168-195-203.us-west-2.compute.internal | Running | 0        | 3 hours    |
| esgf-solr-slave-6b84b5cdb9-5n29d   | ip-192-168-157-180.us-west-2.compute.internal | Running | 0        | 3 hours    |

AWS EKS

Kubernetes clusters

CREATE CLUSTER DEPLOY REFRESH DELETE

A Kubernetes cluster is a managed group of uniform VM instances for running Kubernetes. [Learn more](#)

Filter by label or name

| Name         | Location      | Cluster size | Total cores | Total memory | Notifications | Labels |
|--------------|---------------|--------------|-------------|--------------|---------------|--------|
| esgf-cluster | us-central1-f | 5            | 5 vCPUs     | 18.75 GB     |               |        |

Google Cloud Platform

Kubernetes Engine

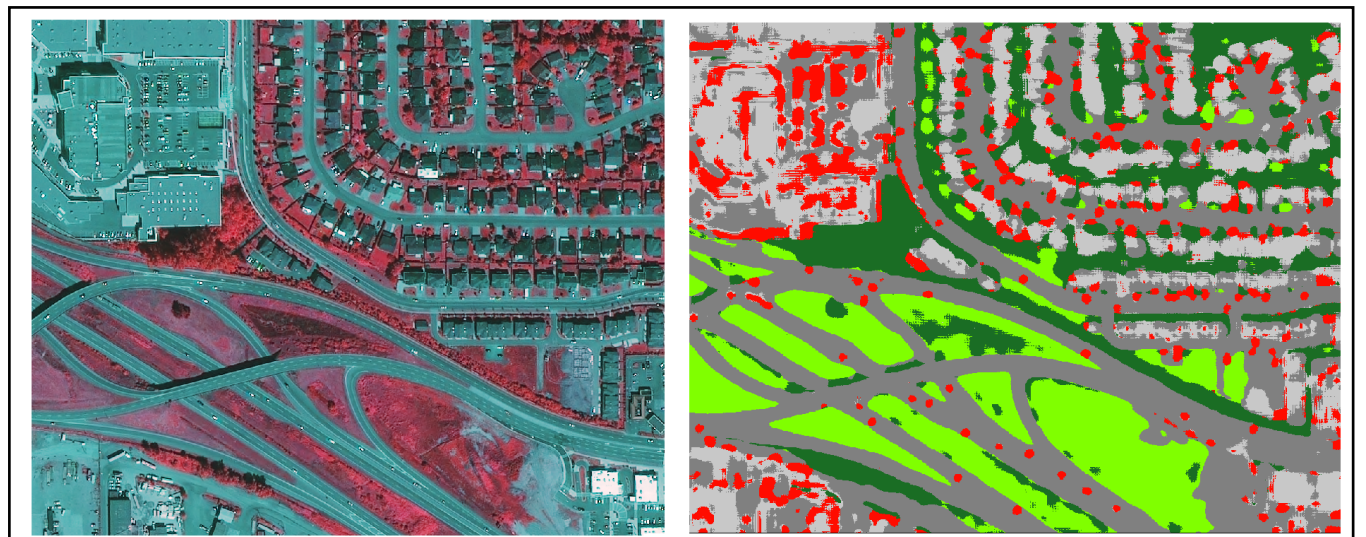
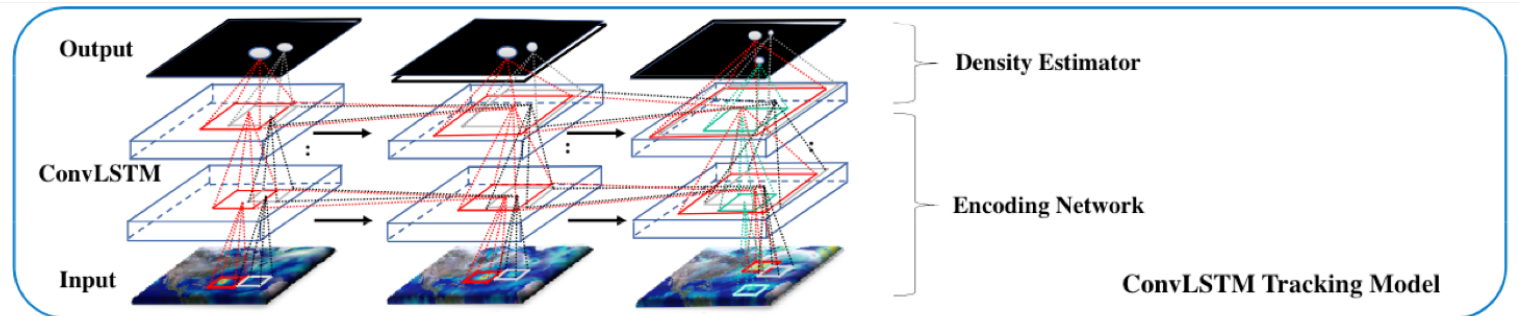
Workloads

| Name               | Status | Type       | Pods | Namespace | Cluster      |
|--------------------|--------|------------|------|-----------|--------------|
| esgf-auth          | OK     | Deployment | 1/1  | default   | esgf-cluster |
| esgf-cog           | OK     | Deployment | 1/1  | default   | esgf-cluster |
| esgf-ids-node      | OK     | Deployment | 1/1  | default   | esgf-cluster |
| esgf-index-node    | OK     | Deployment | 1/1  | default   | esgf-cluster |
| esgf-orp           | OK     | Deployment | 1/1  | default   | esgf-cluster |
| esgf-postgres-auth | OK     | Deployment | 1/1  | default   | esgf-cluster |
| esgf-postgres-cog  | OK     | Deployment | 1/1  | default   | esgf-cluster |
| esgf-postgres-esgf | OK     | Deployment | 1/1  | default   | esgf-cluster |
| esgf-postgres-slcs | OK     | Deployment | 1/1  | default   | esgf-cluster |
| esgf-proxy         | OK     | Deployment | 1/1  | default   | esgf-cluster |
| esgf-slcs          | OK     | Deployment | 1/1  | default   | esgf-cluster |
| esgf-solr-master   | OK     | Deployment | 1/1  | default   | esgf-cluster |
| esgf-solr-slave    | OK     | Deployment | 1/1  | default   | esgf-cluster |
| esgf-ids           | OK     | Deployment | 1/1  | default   | esgf-cluster |

GCP GKE

# Machine Learning

- \* How to mine the vast amounts of data held by ESGF to make reasonable predictions on future global climate and weather events?
- \* LLNL: “Deep Hurricane Tracker” model analyzes patterns in climate simulation data to predict hurricane tracks (S. Kim)
- \* CCMC: High Performance Data Analytics and Machine Learning using Ophidia - an infrastructure for executing declarative, parallel, server side analytics workflows (S. Fiore)
- \* CRIM: working with OGC to advance ML&DL capabilities for high resolution satellite images (T. Landry)





# KEY CHALLENGES



# Key Challenges for 2019 and Beyond

---

- \* Scalability

- \* A container based architecture is highly scalable, but the application itself must be scalable
- \* Must address scalability for publishing services, data catalogs (TDS), and search

- \* Data access

- \* Making it easier for users to download data
- \* Improve or replace the wget scripts
- \* Better support for subsetting at the source, over space and time aggregations

- \* Server side distributed computing

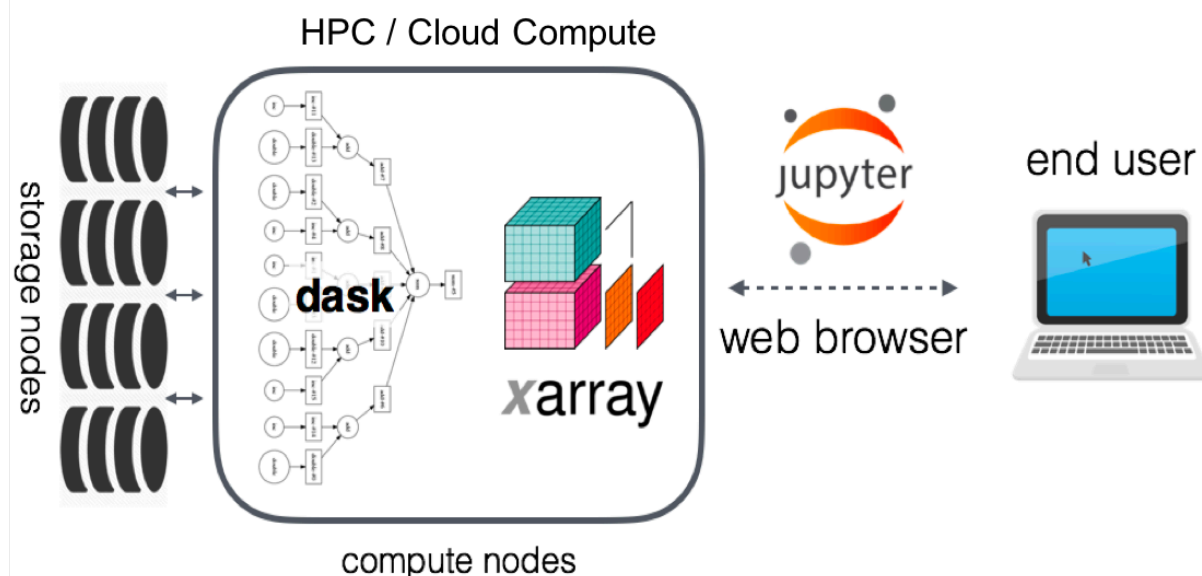
- \* Moving the computation to the data, necessitated by large Big Data volumes
- \* Deploy the Compute Node operationally
- \* Enable workflows that span multiple steps, at multiple sites



# Key Challenges for 2019 and Beyond

## \* Interoperability

- \* ESGF must start to interoperate with other large Earth Science infrastructures
  - \* NASA DAACs, also pushing towards the Cloud
  - \* ESA and Copernicus services (Matt Pryor)
  - \* Pangeo: technical and scientific community building a Python based infrastructure for scalable analysis of Earth Science data (Ryan Abernathy)
    - \* Jupyter for interactive analysis, Xarray/Pandas as data models, Dask for parallel computing
- \* Google dataset search



# Key Challenges for 2019 and Beyond

---

- \* Usability

- \* Re-design the CoG UI to remove some un-needed functionality
- \* Mention the main topics coming up in email support requests
- \* Possibly establishing and supporting a help desk



# Conference Highlights

---

- \* Day 1:

- \* Keynotes and general directions from funding agencies and major stakeholders
- \* Computing services and Machine Learning

- \* Day 2:

- \* ESGF support for CMIP6
- \* Interoperability with other Earth System infrastructures

- \* Day 3:

- \* Reports from ESGF working groups
- \* Open forum on re-thinking the ESGF architecture
- \* Awards Ceremony

- \* Day 4:

- \* ESGF-XC meeting & working groups parallel sessions
- \* Conference summary

- \* Poster session and live demos at the end of each day



